R-ADAPTATION OF THE COURSE “ECONOMETRIC METHODS IN ECONOMICS AND FINANCE”

L. Escot, A. Pérez Alonso, J.E. Sandubete

Faculty of Statistical Studies, Complutense University of Madrid (SPAIN)

Abstract

The course “Econometric Methods in Economics and Finance” includes issues related to data tracking, collection, evaluation and analysis. We adapt the practice sessions in the classroom and the hands-on exercises of this course to the free software $\texttt{R}$; moving from the commercial software $\texttt{EViews}$. We also integrate “DataCamp for the Classroom”, which is a commercial interactive learning platform for Data Science that contains this free module for Academics, into our learning platform CVUCM-Moodle 3.4. We propose new approaches for teaching, learning and assessment in face-to-face and blended environments to improve the e-learning experience of our students. In particular, we prepare teaching and learning resources using $\texttt{R Markdown}$. These resources support self-oriented learning, which would allow us to move to a flipped classroom model. We also develop a specific $\texttt{R}$ package for the course. We use $\texttt{Kahoot}$ and $\texttt{R/exams}$ for summative assessment. The e-assessment is used in conjunction to hands-on exercises to reduce the time taken to mark coursework.

Keywords: R-adaptation, e-learning, hands-on exercises, e-assessment, flipped classroom, self-learning.

1 INTRODUCTION

The course “Econometric Methods in Economics and Finance” was incorporated into the Applied Statistics degree of the Complutense University of Madrid (UCM form its Spanish acronym) as a result of the creation of new study plans adapted to the European Higher Education Area. The objectives of the course were designed taking into account a particular student profile, which is characterized by an interest in developing quantitative and qualitative data analysis skills. These objectives can be summarised in three main points:

- Introduction to the most relevant Economic data sources, both national and international: the National Statistics Institute (INE), the Bank of Spain, the Ministry of Economy and Business (MINECO), the European Statistical Office (Eurostat), the International Monetary Fund, the World Bank, etc.

- Identification of the most appropriate econometric techniques depending on the characteristics of the economic phenomena studied, the data at hand (cross-sectional, longitudinal and time series) and the properties of the explanatory and explained variables analysed.

- Applied computer sessions where students learn to implement the different econometric techniques presented.

The course is raised from an eminently practical approach. We began with the theoretical framework of each chapter (assumptions, estimation and inference) to then move to the practical application based on real data examples.

Previously, the hands-on exercises were taught with Eviews, whose main advantage is that it adapts very well to the content of a basic course in Econometrics. However, as it is a commercial software, its main disadvantage is the cost, both for the UCM and for the students. Moreover, the trend of recent years continues, Eviews does not appear in the rankings of the statistical software most used by the companies that are potentially employers of our graduates in Applied Statistics. On the contrary, Python and $\texttt{R}$, in open source, and SPSS, SAS, Stata and MATLAB are among the statistical programs most demanded by these companies (Muenchen, 2017). This was a clear disadvantage for our students, and thus the primary reason behind moving from the use of the commercial software Eviews to the free software R ($\texttt{r-project.org}$).
2 MOTIVATION

R has a great diversity of libraries and functions. This advantage could turn into a disadvantage for our course, because there is often more than one function that does exactly the same thing. Despite most of the estimation techniques and tests implemented in the course are available in R, they are not integrated into a single package that supports the whole course. R is difficult to use in this regard. In addition, there is no package that includes either the most useful options in R to conduct applied econometric studies or the various options for a specific analysis. The learning method is the opposite of commercial packages such as SPSS or Eviews, where you can explore the menus and the help documentation without knowing all the techniques.

The R-adaptation of the course requires to provide the student with access to a collection of econometric techniques to fill the gap. The starting point of this teaching innovation project is to prepare an easy-to-use index of R libraries and functions that are adapted to the procedures, methods and tests introduced in each lesson.

The problem with this type of listings is their rapid obsolescence because R is a programming language that is constantly updated, self-learning is important. Thus, it is also necessary to provide the students with practice opportunities to play an active role themselves in seeking the functions that are most appropriate for each analysis; or to ask for help to find the solution to specific problems that may arise when working with real world data (R's CRAN Task View, or blogs such as r-bloggers or stack overflow may be very useful for students). These skills are useful for professional competence.

3 OBJETIVES

The general objectives of this teaching innovation proposal are four as follows:

- To adapt the content of the course “Econometric Methods in Economics and Finance” to the statistical software R; Identification of databases from sources of economic information such as the INE, Eurostat, Bank of Spain, MINECO; Look for the most appropriate libraries and functions to implement the techniques explained in each lesson; Implementation of these techniques to real data sets.

  In this sense, we think it is very important to produce written materials tailored for students new to econometrics as well as to prepare an index of libraries and a set of functions for applying the econometric analysis techniques covered in each module.

- To integrate the resources available in the commercial interactive learning platform Datacamp into the Moodle platform of the UCM, CVUCM-Moodle 3.4. In particular, this is done through the DataCamp for the Classroom application, which allows us to assign tasks to student within the platform and incorporate them into the virtual campus.

- To prepare comprehensive teaching and learning resources to give as self-learning material related to the theoretical concepts more relevant to the subject as well as a short introduction to R.

- To apply gamification techniques and other active learning methodologies in education. In particular, we use two tools for student self-evaluation. The Kahoot application is face-to-face, it takes place in the classroom, and it serves as an instrument to review the most important concepts of each lesson; and the other is R/exams library for online self-evaluation. To sum up, we propose new approaches for teaching, learning and assessment in face-to-face and blended environments to improve the e-learning experience of our students.

4 METHODOLOGY

A methodology to reach the four objectives of the R-adaptation of the course is proposed.

- Preparation of induction training materials for R beginners.

- Incorporation of DataCamp for the Classroom.
• Elaboration of learning resources: Data manipulation and exploratory data analysis in R
• Preparation of teaching resources: How to program functions and build libraries in R.
• Adaptation of the course content to R: Defining the libraries and the specific functions required to perform econometric analysis.
• Elaboration of summaries and presentations about the theoretical framework of each module.
• Preparation of documents in R Markdown with practical applications in R for each course module.
  o Heteroscedasticity;
  o Autocorrelation;
  o Unit roots and co-integration;
  o GARCH models;
  o Endogeneity and two-stage least squares;
  o Panel data;
  o Models with qualitative dependent variable;
  o Quantile regression;
  o Causal inference for public policy evaluation;
  o Spatial econometrics.
• Incorporation of resources for self-evaluation in class (Kahoot).
• Elaboration of resources for online self-evaluation in R (R/exams).

5 RESULTS
This section presents our major achievements and describes some pending issues related to the implementation of this innovation project.

• We integrated DataCamp in our course from the application of DataCamp for the Classroom. It is the instructor who must register the students on the DataCamp platform. This module is free for academic institutions and gives premium access to all DataCamp contents for 6 months. As instructors, we can assign tasks and recommend courses or specific lessons of a course among those offered by DataCamp. In addition, the platform allows us to track the students’ progress in each activity. Fig. 1 shows a practical example about the use of DataCamp for the Classroom in our course.

• We continue to explore the possibility of integrating DataCamp for the Classroom into our learning platform CVUCM-Moodle 3.4. Although possible from a technical viewpoint, it requires the approval and support from the IT services of the UCM. Our long-term objective is to centralize all the resources made available to students in their learning process in CVUCM-Moodle 3.4: teaching material of the course, disposable communication channels, to the various options for evaluating the follow-up of the course, and access to DataCamp resources for R self-learning.
• We prepare teaching and learning resources using R Markdown, which is an application available on RStudio. R markdown is a format for writing reproducible documents with R. It offers the possibility of producing a document in PDF, HTML or Word with R code embedded in it. The document combines three elements: explanations in text format for a better understanding of what the R code tries to do in each case; the R code itself with the set of functions used; the results that should be obtained in each practical exercise. We strongly believe that the use of the R Markdown is more didactic than a simple script that only contains the R code of each practice.

Fig. 2 shows a script in HTML format that contains R's introductory training material for beginners. The student should become familiar with the interface and the programming logic of this software, as it will be used throughout the course. In this example, we show how to run a program in R and save the results in different output formats.

Moreover, R Markdown allows us to create eBook format files. We are using this format to write a manual containing practice tasks to reinforce the understanding of the econometric techniques taught in lectures and empirical exercises to illustrate how to apply these techniques in real data. This manual can become an additional helpful reference guide for practitioners (see Heiss, 2016).

As previously mentioned, although R has all the required functions to implement the econometrics methods covered by the course, these functions are not unified in a single library. R Markdown has proven to be a very useful tool for helping us to compile and maintain in the future a detailed compendium of functions with the corresponding R libraries for each chapter of the course.
Fig. 3 shows part of a script in R markdown that was used to explain the problem of omitted variables in a model with cross-section data. After examination of the code, the student must be capable of estimating, in a robust way, the performance of education for married women taking into account the theoretical problem of omitted variables that was detected after diagnosis of the model.
Practical exercises are proposed as part of the continuous evaluation process of the course. Fig. 4 shows a practice that relates back to what we explained about the estimation of the performance of education for married women in Fig. 3. The script integrates the code written in R, the outputs obtained and the comments to explain these results. In this exercise, students are asked to estimate the returns on education. If they suspect that heteroscedasticity may be present, they have to apply the Breusch-Pagan test. If the null hypothesis of homoscedasticity is rejected, a robust estimator for the variance-covariance matrix is required, e.g., White’s robust estimator. To carry out a correct testing and estimation, they have to look for the most appropriate libraries and functions to implement them in R. Practical sessions offer the student the opportunity to play an active role in developing analytical skills that are useful for professional competence, so they should be reinforced.
Ejemplo 1 (Rendimientos de la Educación: Salarios de las mujeres casadas)

```r
modelo.lm <- lm(lwage ~ educ + exper + expersq, data = mroz, subset = (inlf == 1))
summary(modelo.lm)
```

# Summary:
# R Markdown: `summary(modelo.lm)`

```r
# Coefficients:
# Estimate Std. Error t value Pr(>|t|)
# (Intercept) -0.5220405 0.1985321 -2.628 0.00820 **
# educ 0.1074996 0.0164455 7.598 1.94e-13 ***
# exper 0.0415665 0.0131772 3.155 0.00172 **
# expersq 0.0003812 0.0003932 0.766 0.44371 *
# ---
# Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
# Residual standard error: 0.6654 on 424 degrees of freedom
# Multiple R-squared: 0.1568, Adjusted R-squared: 0.1509
# F-statistic: 26.29 on 3 and 424 DF, p-value: 1.30e-15
```

También podríamos haber seleccionado previamente la submuestra de mujeres que participan en el mercado laboral.

```r
misdatos1 <- subset(mroz, is.na(lwage))
modelo1.lm <- lm(lwage ~ educ + exper + expersq, data = misdatos1)
summary(modelo1.lm)
```

```r
misdatos2 <- mroz[complete.cases(mroz),]
modelo2.lm <- lm(lwage ~ educ + exper + expersq, data = misdatos2)
summary(modelo2.lm)
```

Si sospechamos que nuestro modelo puede presentar heterocedasticidad, realizamos el contraste de Breusch-Pagan. La hipótesis nula es H0: Homocedasticidad. La librería `lmtest` incluye el comando `bptest` que permite realizar este contraste pero solo sobre las variables originales del modelo (no las transformadas).

```r
if(!require(lmtest)){install.packages("lmtest")}
library(lmtest)
bptest(modelo.lm)
```

```r
# studentized Breusch-Pagan test
# data:  modelo.lm
# BP = 11.709, df = 3, p-value = 0.008449
```

Como rechazamos la hipótesis nula de homocedasticidad, necesitamos un estimador robusto de la matriz de varianzas-covarianzas. Para calcular errores estándar robustos a heterocedasticidad podemos utilizar los comandos `hcc` (paquete car) o `vcovHC` (paquete sandwich). Si queremos obtener el estimador robusto de White (1980):

```r
if(!require(sandwich)){install.packages("sandwich")}
library(sandwich)
bptest(modelo.lm, vcov=sandwich)
```

---

*Figure 4. Snapshot about the use of R Markdown in our course*
We use two new online applications for summative assessment that provide free and open access to education: Kahoot and R/exams. Kahoot is a game-based learning platform used to host quizzes in class, and R/exams allows for automatic generation and evaluation of exams based on R exercises. Both tools are very effective to ensure the self-learning of the course “Econometric Methods in Economics and Finance”. E-assessment is used in conjunction to hands-on exercises to reduce the time taken to mark the coursework. Fig. 5 shows a couple of examples about the use of Kahoot in our course through an image (above) and a video (below). Students answer different questions using the computer or the mobile phone at the end of each session. This enables us to quickly identify the students who assimilated the concepts from the students who did not. Thus, if it emerges that there are concepts that need clarification, we can always return to them.

6 CONCLUSION
To measure the overall success of the R-adaptation of the course “Econometric Methods in Economics and Finance”, we establish three targets:
• Cost savings derived from the use of free software.

• The new training increases the employability skills and opportunities of our students in the market. Companies, who require analytical workers, select their personnel from people with programming proficiency. Our students acquire and reinforce their skills in the use of the statistical language R, whose use has become more widespread among these potential employers.

• The gradual change in the teaching and learning methods used fosters the transition towards the flipped classroom. This concept was first introduced by Lage, Platt and Treglia (2000) and Bergmann and Sams (2012). It refers to a type of blended learning that combines online material with traditional face-to-face classes. In flipped classrooms, for instance, the students watch online lectures outside of the classroom and take the class time to reinforce the assimilation of contents with the guidance of an instructor and computer-mediated activities. The presence of the instructor is also crucial to motivate and involve the student in her or his learning. Both the didactic resources of DataCamp and the specific materials developed for our course pursue as an ultimate goal the self-learning of the student. We want to gradually replace the classical master class with a mixed method where there is previous preparation by the students of the contents of each chapter with all the available resources on the virtual campus. The face-to-face classes are reserved for active application in front of a computer of different tasks and practices.

To conclude, we would like to reiterate the increasing use of the R software not only in academia, but also in the private sector. Therefore, improving the skills of our students in the use of R implies a knowledge transfer that has a direct impact on society. This project should be continued through the creation of open educational resources within the world of virtual teaching, which is increasingly marked-oriented. If we gather the econometric R package we have developed, the preparation not only of presentations and documents in PDF format to reinforce students’ self-learning, but also the elaboration of audio-visual and interactive material, and the rest of didactic tools, we could offer of an online course on Applied Econometrics in R (perhaps on the DataCamp platform).

ACKNOWLEDGEMENTS

This work was supported by the Government of Spain (grant RTI2018-094901-B-I00); the Complutense University of Madrid - Faculty of Statistical Studies (Innova-Teaching Project No. 275) and the Data Analysis in Social and Gender Studies and Equality Policies Research Group (www.ucm.es/aedipi).

REFERENCES


